Judgment, Graininess and Categories

Ilan Yaniv and Dean Foster
University of Chicago

Judgment plays a role in reasoning, decision making, and planning. An appraisal of a house's value may influence the seller's consideration of various bids, an estimate of a distance may influence a tourist's decision to walk or ride a bus, and the estimated arrival time of a guest may change the dinner schedule. Clearly optimal choice in such cases may hinge on having good judgmental estimates (Tversky & Kahneman, 1974; Dawes, Faust, & Meehl, 1989).

How good a judgmental estimate is, depends on how close it is to the "truth". However, this is not the whole story. People often communicate the "fineness of grain" of their judgmental estimates. For example, compare the following:

(a) John has promised to be at a meeting at "5:00 pm". John arrives at 5:15 pm.

(b) Bill has promised to be at a meeting around "fivish". Bill arrives at 5:15 pm.

Each of the estimates, 5:00 and fivish, conveys different graininess ("grain size"). The first estimate is fine-grain, whereas, the second is coarse-grain. Although the difference between the promised and actual arrival times is the same in both cases, John's late arrival appears to have been more significant because he has committed himself to greater precision by giving a fine-grain estimate. Intuitively, it seems from this example that the quality of a judgmental estimate depends on its graininess.

Communication of graininess in judgment seems ubiquitous. People have at their disposal alternative, parallel scales differing in graininess that allow them to pack efficiently grain size information in their estimates. For example, an hour and 60 minutes refer to identical durations, but differ in their graininess, thus people might pick one or the other depending on their intended precision. Whereas an hour represents a broad category (potentially including values such as 55, 59, and 64 min),

the estimate 60 minutes conveys greater precision. Other pairs such as three weeks vs 21 days and 1/2 a year vs 183 days, to list but a few, involve similar contrasts.

What is the role of graininess in judgment? To begin with, we suggest that the psychological impact of a judgmental error depends on the ratio between the size of the error and the graininess of the judgment. This ratio is called grained error (for definitions see Yaniv & Foster, 1990). To illustrate this concept, consider two hypothetical opinions concerning the date the University of Chicago was founded: (a) Bill's opinion is "in the 1880s" (b) John's opinion is "1885". In light of the correct answer (1892), both opinions are wrong. But the graininess of Bill's estimate is a decade thus his estimate is about "one grain away" from the truth, whereas the graininess of John's estimate is one year, thus his estimate is several (more than one) grains away from the truth.

How do people select the appropriate "grain size" for their judgmental estimates? What role does graininess play in evaluating the accuracy of judgments made by other people? We briefly report here the results of two representative studies (see Yaniv & Foster for full report). In the following section, we outline the paradigm of the first study, our hypotheses and the results.

Grain-Scale study

Method. We asked a group of 44 University of Chicago students to estimate a variety of quantities, such as, "The US population in 1987," "Air distance between Chicago and New York," and " Number of American symphony orchestras." The sample question shown in Figure 1, "The date the University of Chicago was founded," illustrates the format of the questionnaire. Various scales were provided in increasing order of fineness of grain. The top scale was provided as an option in case the respondents didn't know anything about the subject of the question in which case they were supposed circle the whole range. However, if they felt they knew more, they could select finer grain scales. They could circle an interval representing a century on

the second scale (grain size = a century), or else, if they felt they knew the answer with more precision, they could circle an interval representing a fifty-year period on the third scale, etc. They were also given the option of writing down the exact year in the space provided above the sixth scale (grain size = 1 year). Thus, they were supposed to make only one grain-scale estimate per question by selecting an appropriate scale and an interval on that scale. This method jointly elicited their best guesses (approximately the midpoint of the selected interval) along with the graininess of their judgments (represented by the width of this interval).

As part of the analysis, we calculated the grained error of each estimate. For illustration, the grained error of the answer shown in Figure 1 is +2 because it is 2 units away from the interval containing the correct date. (In contrast, the grained error of the estimate 1900-1909 would have been -1, while the grained error of 1890-1899 would have been 0 because this interval contains the correct date.)

Hypotheses. One hypothesis is that in selecting grain sizes for their judgments, individuals signal to others the magnitude of judgmental error that they expect. Thus, individuals estimate historical dates in decades if they believe their best guess might be (on the average) 10 years off the truth, and they give estimates in centuries if they expect an average (absolute) error on the order of 100 years. A corollary of this hypothesis is that the mean absolute grained error across judgments should average one.

Another possible hypothesis is that grain size of a judgment is like a "confidence interval" -- an interval that includes the true answer with measured certainty, for example, with a probability of 95% or 99%. If this is true, then grain-scale judgments should include the correct answers with a high probability.

Results. The distribution of the grained errors across all grain-scale judgments is plotted in Figure 2. The most striking result is that only 46% of the grain-scale judgments actually contained the correct answer; 75% of the judgments had a grained error that was either -1, 0, or +1; and 95% of the judgments had a grained error that

ranged from -5 to +5. Thus, grain-scale judgments do not appear to "behave" like confidence intervals that contain the correct answers with a high probability. These results are consistent with Alpert and Raiffa's (1982) who found that judgmental confidence intervals tend to be too narrow (see also Yates, 1990).

It is interesting that grain-scale judgments tend to be "too finely-grained" and hence exclude the truth so frequently. Clearly, giving coarsely-grained estimates can reduce grained error and increase the chances that the truth is included in the estimate. For instance, a coarse-grain estimate for the date the University of Chicago was founded "1700-1900" will generate a lower grained error than a fine-grain estimate such as "in the 1880s."

Grained Error vs Graininess: A Tradeoff

In everyday situations, however, giving excessively coarse judgments is discouraged by linguistic/social norms which imply that speakers should be truthful and appropriately informative (Grice, 1975). Judgmental estimates are expected to be truthful, that is, have low grained error. But, they are also expected to be informative, or finely-grained. For instance, the forecast that inflation rate in the US in 1990 will be between 1% and 55% appears vacuous although it is quite likely to be truthful. Consider for example two estimates of the age of a particular person: (a) "25 to 40 years" (b) "40 to 42 years." Suppose that the person's true age is 39 years. In retrospect, which estimate would we prefer to have had? The first estimate is coarser than the second and has a lower grained error. The second estimate is more informative, but it has a higher grained error. The evaluation of estimates appears to involve a tradeoff between grained error and graininess. A model which formalizes this tradeoff between the truthfulness (grained error) and informativeness (graininess) of a given judgment is presented in Yaniv and Foster (1990).

Evaluation Study

We briefly outline a study designed to test our model. We told our respondents to "imagine that you are a senator preparing an argument. You solicit quick judgmental estimates for some missing information from two of your aides. Later, you compare these estimates to the truth." We presented to them pairs of grain-scale estimates of the target questions along with the correct answer. For example:

> Amount of money spent on education by the U.S. federal government in 1987?
> _____ Aide A responds: $20 to 40 billion
> _____ Aide B responds: $18 to 20 billion
> The actual answer was: $22.5 billion. Which aide is more credible?

In each case, respondents were supposed to compare the estimates in terms of their quality, specifically, to indicate which of the two aides they would prefer to consult in the future.

Results. Our model provided a weighted measure of the "quality" of each estimate based on its grained error and graininess (width of the judgmental interval). We found that the preferences implied by our model predicted the preferences indicated by our respondents.

Final Comments

We have discussed two major ideas. First, judges seem to select grain sizes which estimate the expected error of their estimates. Second, the graininess of a given judgment depends on the tradeoff between the judge's pragmatic (and conflicting) needs to be truthful and informative at the same time (for further discussion see Yaniv & Foster).

It would be interesting to speculate on the representation that may give rise to grain-scale judgments. It is conceivable that grain-scale judgments rely on permanent hierarchical memory representations similar to those underlying natural language categories (e.g., Smith & Medin, 1981). Recent work (Huttenlocher et al., 1990)

suggests that the dates of particular autobiographical events (e.g., the date we saw a particular movie) might be coded at multiple levels (e.g., "January 20", "January" and "Winter quarter".) It is possible that numerical information is generally coded in such structures. For example, historical events (e.g., the date the University of Chicago was founded) may be coded at multiple levels varying in fineness (similar to the levels represented in Figure 1) and with different degrees of certainty. Thus we may remember the target event was "definitely a 19th century event", "most likely in second half of the 19th century," "possibly in the 1880s," etc. Grain-scale judgments may result from the confluence of several such sources of information.

References

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky, Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

Dawes, R.M., Faust, F., & Meehl, P.E. (1989). Clinical versus actuarial judgment. Science, 243, 1668-1674.

Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), Syntax and semantics: Vol. 3. Speech acts (pp.64-74). New York: Academic Press.

Huttenlocher, J., Hedges, L., & Bradburn, N. (1990). Reports of elapsed time: Bounding and rounding processes in estimation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 196-213.

Smith, E.E., & Medin, D.L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

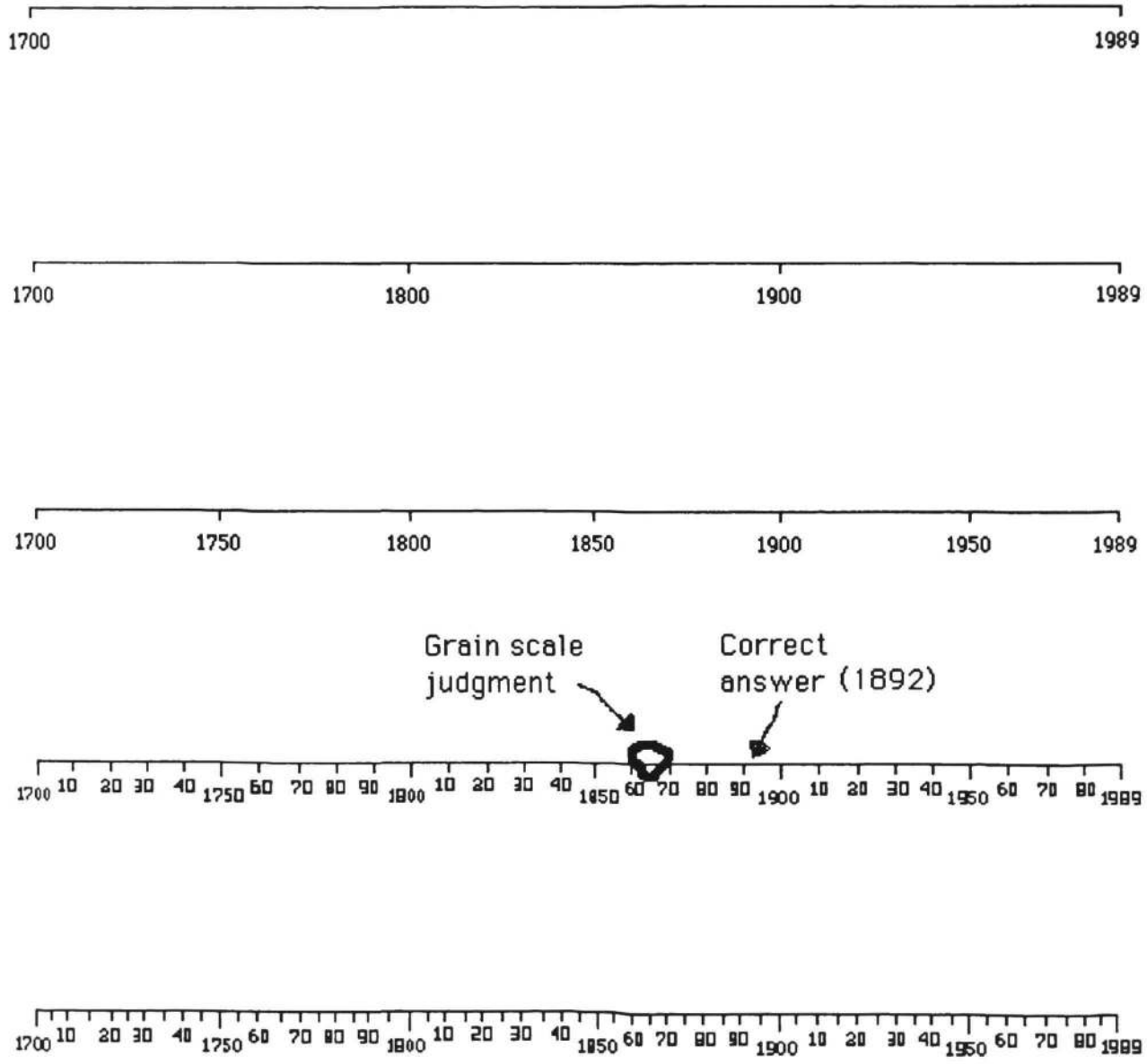
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Yaniv, I., & Foster, D.P. (1990). Graininess of judgment. Center for Decision Research, University of Chicago.

Yates, J.F. (1990). Judgment and Decision Making (pp. 75-111). Englewood Cliffs, NJ: Prentice Hall.

Figure 1.

Date the University of Chicago founded?



The exact year is: _____

Figure 2.

Grain-Scale study: Distribution of Grained Errors